DOCUMENT RESUME

ED 030 582 SP 002 197

By-Abramson, Theodore
Reliability of an Observation of Teachers' Classroom Behavior.
School University Teacher Education Center, N.Y.
Spons Agency-Office of Education (DHEW), Washington, D.C.
Pub Date 69
Contract-OEC-1-6-050945-1673
Note-15p.
EDRS Price MF-$0.25 HC-$0.85
Descriptors-*Analysis of Variance, Check Lists, *Classroom Observation Techniques, *Data Analysis, Instructional Innovation, *Models, Observation, *Reliability, Teacher Education
Identifiers-Observation Schedule, School University Teacher Education Center, SUTEC

Although the evaluation of innovations in teacher training is often dependent on observational data, the problem of the reliability of the observations collected by a team of observers has often been treated in a superficial manner, partially because of the difficulties of maintaining an observer team intact over an extended period of time and of being permitted to observe each teacher a number of times. An analysis of variance model has been developed which permits the calculation of an overall reliability coefficient and the partitioning of the sources of variation for the typical observer team situation in which the team visits a number of different teachers only once and where the team does not necessarily contain the same members for all visits. The paradigm is developed for the situations in which there are "n" observations per item per observer as well as when there is only one observation per item per observer. The model has been tested using data based on the School University Teacher Education Center (SUTEC) observation schedule which is designed to investigate seven aspects of classroom behavior. Since the proposed model permitted the partitioning of the variance associated with the component parts of the schedule, it may provide useful as a test of the homogeneity of the items in an observation schedule as well as for reliability calculations. (Author/JS)

Reliability of an Observation of Teachers' Classroom Behavior[1]

Theodore Abramson

~~Fordham University~~[2] City College of New York

During the last 15 years a number of observational schedules of teachers' classroom behavior have been formulated (Medley & Mitzel, 1963). The reliability of the observer teams using these instruments has often been dealt with superficially.

Recently, Denny (1968) used an analysis of variance (ANOVA) technique to calculate the reliability of an observation schedule. The ANOVA technique used was essentially that first proposed by Medley and Mitzel (1958, 1963) and permitted the partialling out of various sources of variance as well as the calculation of a reliability coefficient. The Medley and Mitzel (1963) technique required that the same observers visit the same teachers a number of times. The difficulties of maintaining a sizable observer team intact for an extended period of time and of observing the same teachers a number of times make the ANOVA model difficult to apply in many typical situations which call for observational data.

This paper presents an ANOVA model which permits partialling out of variances and reliability calculations when the observer team does not have the same observers throughout and when the observation of each teacher occurs only once. It is felt that this model is applicable to many observer team situations, since the typical team is trained by observing the same phenomena as a group and then comparing observations. The model is then applied to the "live" portion of the School University Teacher Education Center (SUTEC) observation team data.

Method

Since each teacher is observed by an observer team peculiar to himself, the model may be considered a partially hierarchical design. That is, each observer team has the same number of observers but not necessarily the same observers, and therefore the observer team factor is nested under the teacher factor. If teachers are factor A, observers factor B, and items factor C, B would be nested under A. Assuming that there are n scores on each item for each teacher per observer the sources of variation, degrees of freedom, and expected mean squares are as given in Table I (Winer, 1962) where p, q, and r are the numbers of teachers, observers, and items respectively.

Insert Table 1 about here

The D_p , D_q , and D_r terms are equal to 1-p/P, 1-q/Q, 1-r/R, respectively, where the p and P, q and Q, and r and R are the sample and population parameters of teachers, observers, and items, respectively. Each of these D's is either 0 or 1 depending on whether the corresponding factor is fixed or random.

As was pointed out by Medley and Mitzel (1963), the assignment of a variable as fixed tends to reduce the error of measurement and hence inflate the reliability and therefore the assumption that a variable is fixed should be based on sound reasons. A rule of thumb for selecting which factors are fixed and which are random is to decide whether other elements comprising the factor might have been used, and if so, then the factor is random (Medley & Mitzel, 1963). For example, if no observers other than the ones actually employed could have been used satisfactorally, then the observer factor would be fixed. Since there are always other teachers and observers available, theoretically anyway, these factors are considered random factors.

More precisely, as p, q, and r, the number of the sample elements, approach the values of P, Q, and R the number of elements in the population, the ratios p/P, q/Q, and r/R approach a value of one and therefore D_p, D_q, and D_r approach zero. If zeros are substituted for the D's the number of factors contained in the expected mean squares shrink and thus the reliability is increased because the denominator of the fraction which defines the reliability coefficient is decreased.

The model is also applicable even when there is only one score per item per observer for each teacher. In this case the model is the same as in Table 1 with n=1 and the within source of variation removed. If all factors are random, ones are substituted for the D's and the model now yields an error term of $\sigma^2_{bc} + \sigma^2_e$ (Winer, 1962). The remaining expected mean square values follow in a similar fashion. To simplify the model still further the Medley and Mitzel (1963) procedure may be utilized. According to this procedure, the last term in the source of variation column, the residual, is considered to be the error term and is denoted by σ^2_e rather than $\sigma^2_{bc} + \sigma^2_e$. The simplification of the error term and the substitution of ones for the n and the D's result in the expected mean squares shown in Table 2.

- - - - - - - - - - - - - - - - - - - -
Insert Table 2 about here
- - - - - - - - - - - - - - - - - - - -

The only major difference between the Winer (1962) and Medley and Mitzel (1963) approach occurs in the F ratio testing the main effects of factor A. This particular F ratio utilizes the nested factor B as its denominator, and has a bigger expected mean square term in the simplified version than is called for by Winer (1962). The difference between the models is due to the σ^2_{bc} term. This therefore means that a significant F ratio testing the hypothesis $\sigma^2_a=0$ in the simplified version would certainly be significant according to Winer (1962). Since the other two F ratios testing the hypotheses $\sigma^2_c=0$ and $\sigma^2_{ac}=0$ use the residual expected mean square as denominators, both the Medley and Mitzel (1963) and Winer (1962) approaches yield the same F values in these two cases.

There are actually two homogeneity assumptions implied by the model. The first is that the source of variation due to B(A) represents the pooled variation of observers within teachers. The second results from the fact that the residual term is actually the B(A)XC interaction term and represents the pooling of different sources of variations. The homogeneity assumption here is equivalent to the assumption that the correlation between items is constant within each of the teachers.

The model was applied to an observation schedule which was developed by a research team at a teacher training institute to investigate certain aspects of the classroom behavior of the institution's first year graduates who were teaching in the New York City public school system. The observer team was to observe only the following seven categories of behavior: Teacher mobility, involvement of children, materials present, materials in use, directed behavior, spontaneous behavior, and irrelevant acts. These items are briefly described below. More detailed descriptions are available (Chapline, 1968).

Teacher mobility. The number of different positions occupied by the teacher during the second five minutes of each learning activity--indicated on a room sketch.

Involvement of children. A global judgement of the attentiveness of the whole class during each learning activity--assessed on a three point scale from uninvolved (1) to highly involved (3).

Materials present. The number of different materials present during the entire observation--checked on a list of materials.

Materials in use. The number of different materials in use during the entire observation--checked on a list of materials.

Directed behavior. The number of times during each activity that the teacher called on pupils without the pupils first indicating a willingness to respond.

Spontaneous behavior. The number of times that the pupils indicated a willingness to respond before being asked to do so plus the number of times that the pupils responded spontaneously before permission was granted. The score on this category was weighted in a ratio of 1:2, respectively, before being added. Raising hand behavior would be scored as a one while calling out the answer would be scored as a two. If both occurred during the same activity, the activity would be scored as a three provided nothing else happened for the duration of the activity.

Irrelevant acts. The number of behaviors or movements obviously not related to the learning activity of 12 randomly selected children. Each child was intensively observed for a two minute period.

Three teachers were observed once through a one way glass by three different observer teams. Each observer team contained seven members, but some of the observers were not the same throughout all the observations and therefore the teams were considered different.

In line with the earlier discussion of random and fixed variables, the teacher and observer factors were considered random factors, but because the observers were instructed to disregard all behavior other than those on the observation schedule the items were considered fixed. Accordingly, the σ^2_{ac} term in the first and third lines of Table 2 were dropped from the expected mean squares for teachers and items, respectively. The actual and expected mean squares for this specific situation in which p=3, q=7, and r=7 are given in Table 3.

Insert Table 3 about here

The notations for the observed mean squares used in Table 3 and the symbol "(=)" (to be read "is estimated by") in Table 4 come from Medley and Mitzel (1958).

The general set of linear equations which must be solved to find the estimated variance components is constructed by setting the estimated mean square terms equal to their corresponding observed mean squares. The resulting linear equations are then solved simultaneously.

Table 4 gives the particular set of linear equations for the specific case listed in Table 3 and the resulting estimated values of the variances for each factor.

Insert Table 4 about here

Results

The three hypotheses $\sigma_a^2=0, \sigma_c^2=0, \sigma_{ac}^2=0$ were all rejected because their respective F ratios,

$$F_a = \frac{MS_a}{MS_{b(a)}} = 6.3124,$$

$$F_c = \frac{MS_c}{MS_{residual}} = 117.3776,$$

$$F_{ac} = \frac{MS_{ac}}{MS_{residual}} = 19.4667,$$

~~and~~ were all significant at the .01 level. The appropriate df's are given in Table 3. The rejection of these three hypotheses indicated that the scale does differentiate between teachers and items, and that there is a significant interaction between these two non nested factors.

The overall reliability coefficient (Medley and Mitzel, 1963) is equal to $R_{xx} = \frac{\sigma_T^2}{\sigma_x^2}$. Here $\sigma_T^2 = (qr)^2 \sigma_a^2 = (7\cdot 7)^2 \sigma_a^2 = 49^2(.7222)$

$$\sigma_T^2 = 1734.0022$$

and $\sigma_x^2 = qr(qr\,\sigma_a^2 + r\sigma_{b(a)}^2 + q\sigma_{ac}^2 + \sigma^2)$

$$\sigma_x^2 = (7\cdot 7)\left[(7\cdot 7)\ (.7222)+7(.5376)+7(7.6460)+2.8983\right]$$

$$\sigma_x^2 = 4682.9937$$

Therefore, $R_{xx} = \frac{1734.0022}{4682.9937} = .37.\cancel{02\%}$

The .37 reliability coefficient indicated that 37% of the variance was attributable to the teacher factor and 63% of the variance was due to the items, interaction, and residual factors. An examination of the ratio of the variances due to teachers and observers, the factor nested under teachers, indicated that 21.2% and 15.8% of the component of the total variance due to teachers was due to teachers and observers, respect-ively. A similar calculation for the other factors comprising the remaining 63% of the total variance yielded values of 38.0%, 18.1% and 6.9% for the items, interaction, and error or residual terms respectively.

Discussion

The proposed model did permit the partitioning of the variance associated with an observational schedule into its component parts and the calculation of an over all reliability coefficient. In the particular case to which the model was applied 75% of the variance was due to teachers and items, each of these two factors contributing equally to the total variance. Only 15.8% of the total variance was due to observers; the factor nested under teachers. These facts permit one to conclude that the variance due to different observers being used was considerably smaller than that due to the different teachers as they were observed on the various types of behavior represented by the items of the observational schedule.

That the items accounted for the single largest source of variance was probably due to the very different elements of behavior being observed. For example, materials present required very little judgement on the part of the observer, while involvement of children required a great deal of judgement. Indeed, one of the proposed future uses of the paradigm presented is as a test of the homogeneity of the items in observational schedules and therefore the model may be found useful when applied to data based on other observational instruments for more than reliability calculations.

Footnote 1. The project reported herein was supported by the United States Department of Health, Education, and Welfare, Office of Education Contract No. OEC 1-6-050945-1673 and was conducted at the School University Teacher Education Center.

Footnote 2. Graduate research training fellow under the New York State Cooperative Research Training Program.
Currently at the City College of New York.

References

Chapline, Elaine. School University Technical Progress Report. September 1968, Cooperative Research Project no. 5-0945, contract no. OEC 1-6-050945-1673, United States Office of Education.

Denny, David A. "Identification of Teacher-Classroom Variables Facilitating Pupil Creative Growth," American Educational Research Journal 5:365-383; May 1968.

Medley, Donald M. and Mitzel, Harold E. "Application of Analysis of Variance to the Estimation of the Reliability of Observations of Teachers' Classroom Behavior," Journal of Experimental Education 27: 23-35: September 1958.

Medley, Donald M. and Mitzel, Harold E. "Measuring Classroom Behavior by Systematic Observations" in N.L. Gage (Ed.), Handbook of Research on Teaching Chicago: Rand-McNally Co., 1963, pp. 247-328.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962, 672 pp.

TABLE I

Sources of Variation, Degrees of Freedom, and Expected Mean Squares for an ANOVA Design with Factor B Nested Under Factor A

Source of Variation	d f	E (MS)
A	p-1	$nqr\sigma_a^2 + nrD_q\sigma_b^2 + nqD_r\sigma_{ac}^2 + nD_qD_r\sigma_{bc}^2 + \sigma_e^2$
B W.A	p(q-1)	$nr\sigma_b^2 + nD_r\sigma_{bc}^2 + \sigma_e^2$
C	r-1	$npq\sigma_c^2 + nqD_p\sigma_{ac}^2 + nD_q\sigma_{bc}^2 + \sigma_e^2$
AC	(p-1) (r-1)	$nq\sigma_{ac}^2 + nD_q\sigma_{bc}^2 + \sigma_e^2$
(B W.A) C	p(q-1) (r-1)	$n\sigma_{bc}^2 + \sigma_e^2$
Within	pqr(n-1)	σ_e^2

TABLE 2

ANOVA Design with Factor B Nested Under

Factor A, All Factors Random, and n = 1

Source of Variation	d f	E (MS)
A	p-1	$qr\sigma^2_a + r\sigma^2_{b(a)} + q\sigma^2_{ac} + \sigma^2$
B W.A	p(q-1)	$r\sigma^2_{b(a)} + \sigma^2$
C	r-1	$pq\sigma^2_c + q\sigma^2_{ac} + \sigma^2$
AC	(p-1) (r-1)	$q\sigma^2_{ac} + \sigma^2$
Residual	p(q-1) (r-1)	σ^2

TABLE 3

Analysis of Variance of an Observation Schedule Containing Seven Items and Using Three Observer Teams and Three Teachers

Source of Variation	d f	E(MS)	Observed (MS)
A (Teachers)	2	$49\sigma_a^2 + \sigma_{b(a)}^2 + \sigma^2$	$s_a^2 = 42.0509$
B(A) (Observers within Teachers)	18	$7\sigma_{b(a)}^2 + \sigma^2$	$s_{b(a)}^2 = 6.6616$
C (Items)	6	$21\sigma_c^2 + \sigma^2$	$s_c^2 = 340.1953$
AC	12	$7\sigma_{ac}^2 + \sigma^2$	$s_{ac}^2 = 56.4206$
Residual	108	σ^2	$s^2 = 2.8983$

TABLE 4

Estimation of Variance Components for an Observation Schedule Containing Seven Items and Using Three Observer Teams and Three Teachers

$$\sigma_a^2 \ (=) \ \frac{1}{49} \left(s_a^2 - s_{b(a)}^2 \right) \doteq .7222$$

$$\sigma_{b(a)}^2 \ (=) \ \frac{1}{7} \left(s_{b(a)}^2 - s^2 \right) = .5376$$

$$\sigma_c^2 \ (=) \ \frac{1}{21} \left(s_c^2 - s^2 \right) = 16.0618$$

$$\sigma_{ac}^2 \ (=) \ \frac{1}{7} \left(s_{ac}^2 - s^2 \right) = 7.6460$$

$$\sigma^2 \ (=) \ s^2 = 2.8983$$